# Commentary

Imprinted expression of small nucleolar RNAs in brain: Time for RNomics

Witold Filipowicz*

Friedrich Miescher Institute, P.O. Box 2543, CH-4002 Basel, Switzerland

Biochemistry students of my generation had the RNA story relatively straight: ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), and messenger RNAs (mRNAs) appeared to do everything expected of RNA in the cell at that time. In the last two decades or so, the situation has changed dramatically. In addition to rRNAs and tRNAs, hundreds of other non-protein-coding RNAs have emerged, having a diverse range of functions, from structural through regulatory to catalytic (1, 2). Particularly exciting recent examples of regulatory RNAs include *Xist* and *roX* RNAs, involved in chromosome dosage compensation in mammals and *Drosophila*, respectively (3), and tmRNA, which directs tagging of incomplete proteins for degradation in bacteria (4). By its nature—a potential to base-pair—the recognition of sequences in other nucleic acids is an RNA molecule's major asset. Indeed, several spliceosomal small nuclear (sn) RNAs (e.g., U1, U2, and U6) base-pair with short sequences in pre-mRNAs to delineate regions to be spliced out (5, 6). Other RNAs, which identify their targets by base-pairing, include editing guide RNAs in the mitochondria of kinetoplastid protozoans (7), and *lin-4* and *let-7*, the 21-nucleotide (nt) long mini-RNAs, which regulate timing of early developmental decisions in *Caenorhabditis elegans* by hybridizing to the 3′ untranslated regions of their target mRNAs (8); *let-7* functions also in other bilaterian animals, from molluscs to mammals (9).

The class of noncoding RNAs that—at least in terms of numbers—is currently dominating the field is small nucleolar (sno) RNAs, which act as guides to direct pseudouridylation and 2′-O-ribose methylation in rRNA (10–14). In vertebrates, each of these modifications is found at about 100 sites per ribosome. Since a single guide snoRNA specifies one, at most two, modifications (Fig. 1*A*), the number of these RNAs may approach 200. Guide snoRNAs also operate in other territories; e.g., pseudouridylation-guide-like RNA and associated proteins form part of the mammalian telomerase (refs. 15 and 16 and references therein). Bio-
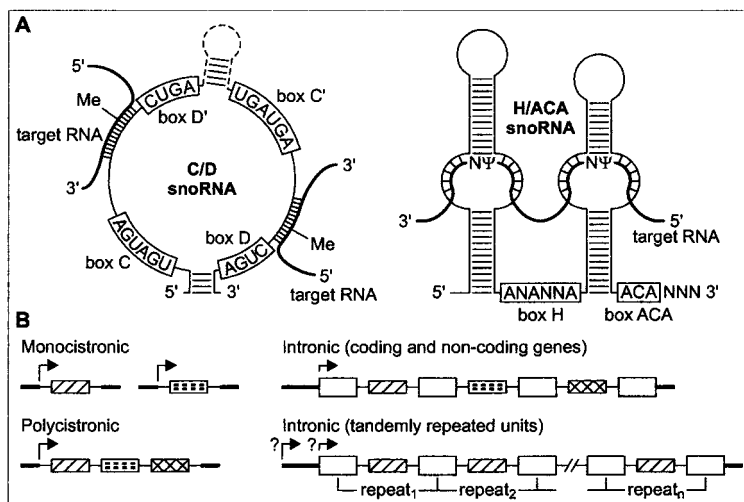
Fig. 1. Short guide to the structure (*A*) and biogenesis (*B*) of guide snoRNAs. (*A*) Schematic structure of the C/D-box and H/ACA-box snoRNAs, with conserved sequence elements and base-paired target RNAs indicated. Most snoRNAs contain one rather than two functional modification domains, although H/ACA-box snoRNAs always have a conserved hairpin-hinge-hairpin-tail secondary structure. The pseudouridylation pocket forms two short duplexes (3–10 bp) with the target RNA, leaving the uridine residue to be isomerized unpaired. snoRNAs of each class are associated with a set of specific proteins which are not shown. (*B*) Different strategies of snoRNA expression. In the yeast *Saccharomyces cerevisiae*, most of the snoRNA genes are transcribed as either mono- or polycistronic units, and only a few are encoded in introns. In all established cases, vertebrate guide snoRNAs originate from introns of either protein-coding or noncoding RNA polymerase II-transcribed genes. Sequences corresponding to mature snoRNAs are shown as filled-in boxes, and exons, as open boxes. Arrows indicate transcription initiation sites. Transcription start sites for the tandemly repeated brain-expressed snoRNA genes are not known.

genesis of guide snoRNAs in different organisms is equally interesting (Fig. 1*B*). In vertebrates, all guides are encoded in introns of genes transcribed by RNA polymerase II (pol II), but some of the host genes do not code for proteins, and it is their introns and not exons that are evolutionarily conserved (10, 11, 14). The paper by Cavaillé *et al.* (17), appearing in this issue of PNAS, adds still another dimension to the guide snoRNA story. These authors characterize a collection of brain-specific snoRNAs whose expression is paternally imprinted. Interestingly, genes encoding most of the identified RNAs map within a region implicated in the neurogenetic disease Prader–Willi syndrome (PWS). One of the imprinted snoRNA loci [MBII-85/HBII-85; nomenclature of Cavaillé *et al.* (17)] is also described by De los Santos *et al.* (18).

The snoRNAs fall into two major classes, each characterized by specific conserved sequence elements ("boxes") and a set of associated proteins (Fig. 1*A*). The C/D box snoRNAs, associated with fibrillarin, guide a site-specific 2′-O-methylation, and the H/ACA box snoRNAs, associated with protein GAR1, target specific conversions of uridine to pseudouridine. Guiding of 2′-O-methylation involves base-pairing of the 10- to 21-nt-long sequence positioned upstream of box D (or D′) to the target RNA, with the nucleotide positioned 5 base pairs (bp) upstream of the D/D′ box being selected for methylation. In the H/ACA snoRNAs, one or both of the two RNA hairpins are

interrupted by an internal loop, the pseudouridylation pocket, which contains two short (3- to 10-nt) sequences complementary to nucleotides flanking the site of isomerization (refs. 10–14; Fig. 1*A*).

When searching for small RNAs expressed in mouse brain, Cavaillé *et al.* (17) identified three novel C/D box snoRNAs (MBII-13, MBII-52, and MBII-85) and one H/ACA box snoRNA, MBI-36. In mouse or rat, all four RNAs are present exclusively in the brain, with the three C/D box RNAs accumulating at similar levels in all brain areas except the choroid plexus, and the H/ACA RNA MBI-36 having the reverse distribution, being found mainly in the choroid plexus. In humans, all mouse RNA orthologs are likewise either exclusively (HBII-52 and HBI-36) or predominantly (HBII-13 and -85) expressed in brain.

The most interesting aspect of the work is that the genes encoding all three human C/D box snoRNAs map to chromosome 15q11–q13, more precisely to the ≈1.5-Mb region linked to PWS; they also map to the syntenic chromosome 7C region in mouse (17, 18). Genes in the PWS region are parentally imprinted, with only alleles inherited from the father being expressed (19–23). The loss of imprinted gene expression, most frequently caused by paternal deletions or maternal uniparental disomy, results in severe developmental and neurobehavioral problems. PWS occurs in about 1 in 15,000 births. Its clinical features include infant failure to thrive with hypotonia, hyperphagia leading to severe obesity, hypogonadism, and mild to medium mental retardation with learning and other disabilities. Among the paternally controlled genes in the PWS region are some that encode proteins—e.g., the dicistronic gene *SNURF-SNRPN*, coding for the brain-expressed protein SmN, a counterpart of the common U-snRNP Sm proteins B and B'. Other regions produce RNAs without apparent protein-coding potential, and some of the noncoding transcripts are antisense to translatable mRNAs, a situation frequently encountered in different parentally imprinted regions (20, 21, 23).

Consistent with their chromosomal localization, the C/D box snoRNA genes are paternally imprinted and are not expressed in PWS patients or mouse models of the disease (17, 18). Detailed analysis of the imprinted snoRNA loci revealed their very unusual organization. snoRNA sequences of one type (MBII-52/HBII-52) are embedded in ≈2-kb DNA units which, in humans, are tandemly repeated 47 times (17); snoRNAs appear to be processed from the excised introns by the exonucleolytic mechanism (24). The MBII-85/HBII-85 snoRNAs are likewise present in multiple cotranscribed repeats,

but they are less uniformly spaced and their mode of processing is not certain (17, 18). Sequences flanking the snoRNA regions are not phylogenetically conserved and have no protein-coding potential, making the host genes of these brain-expressed snoRNAs similar to the noncoding hosts of other snoRNAs (Fig. 1*B*).

The organization and expression of other genes in the PWS region is also very complex. In addition to snoRNA loci, many other DNA repeats are present; it is conceivable that DNA repeats contribute to the establishment of imprinted states. Expression of all genes in the PWS region, and also maternally imprinted genes distal to it and affected in Angelman syndrome (AS, another neurogenetic disease), is controlled by the large bipartite imprinting center (IC). The IC region partially overlaps the upstream portion of *SNURF-SNRPN*, making functional analysis of individual genes difficult. Current models propose that loss of expression of multiple paternal-specific genes contributes to the PWS phenotype. At least two of the described snoRNA loci map to the regions implicated in the disease. It is likely, however, that other paternally imprinted genes will be found in the many still-uncharacterized regions of 15q11–q13. PWS and AS phenotypes strongly argue for a function of imprinting in brain development. Furthermore, the expression data for 15q11–q13, and also other imprinted loci, indicate that noncoding RNAs may play an important role, either in imprinting itself or in transmission of imprinted information (19–23).

What could be the biological role of the newly described brain-specific snoRNAs? Lack of complementarity to rRNA and their tissue-specific expression argue against the possibility that they specify rRNA modification. Significantly, recent work has shown that guide snoRNAs can also function in posttranscriptional modification of cellular RNAs other then rRNA. In vertebrates, addition of several 2′-*O*-methyls to the spliceosomal U6 snRNA is directed by classical snoRNAs (25, 26), and the 2′-*O*-methylation and pseudouridylation at two adjacent positions in U5 snRNA are both guided by a novel C/D-H/ACA-box "hybrid" RNA (T. Kiss, personal communication). Identification of "orphan" snoRNAs for which targets are unknown (27), and the demonstration that snoRNAs can, although inefficiently, modify sequences inserted into RNA polymerase (pol) II-transcribed mRNA-like molecules (28), further argue for substrate diversity. The small C/D-box-like RNAs present in Archaea direct methylation of not only rRNAs but also tRNAs (refs. 29 and 30; P. Dennis, personal communication; J.-P. Bachellerie and C. Gaspin, personal communication).

The three brain C/D box snoRNAs show no significant potential to base-pair with known cellular small RNAs but one of them, MBII-52/HBII-52, has a guide region with an 18-nt phylogenetically conserved complementarity to the serotonin receptor 5-HT$_{2C}$ mRNA that is expressed in brain (17). Interestingly, a potential acceptor of the methyl group, an A residue, happens to be the subject of the physiologically relevant A-to-I editing in this mRNA (31, 32). Because 2′-*O*-methylation of adenosine in a model substrate is known to decrease its deamination by 200-fold (33), regulation of mRNA editing by snoRNAs is an attractive possibility, particularly in brain. The A-to-I editing enzymes are most abundant in brain (34, 35), and quantification of the inosine content in poly(A)$^+$ RNA suggests that hundreds of mRNAs might undergo editing in this tissue (34). If the snoRNA-directed 2′-*O*-methylation of mRNA indeed occurs *in vivo* (although 2′-*O*-methyls have been found only in 5′-cap-proximal nucleotides of mRNA to date; ref. 36), editing might not be the only process potentially regulated by such modification. The 2′-*O*-methylation of 3′ splice site and polyadenylation signal (AAUAAA) nucleotides strongly inhibits the respective pre-mRNA processing reactions *in vitro* (37, 38). One could also envisage effects of base-pairing of snoRNAs to mRNAs (or other RNAs), independently of methylation. Such associations, related to the proposed chaperonin function of snoRNAs in ribosome biogenesis (10, 14), could modify maturation patterns of target RNAs by masking the processing signals or interfering with binding of protein factors. Interestingly, the snoRNA complementarity in 5-HT$_{2C}$ mRNA occurs near a site used for alternative splicing (17).

How could guide snoRNAs, localized predominantly in the nucleolus, target nucleoplasmic substrates such as snRNAs or mRNAs? One possibility is that they meet within Cajal (coiled) bodies (CBs). Nearly all components of C/D box and H/ACA box snoRNAs, and also U snRNPs, are present in CBs of mammalian cells (refs. 16 and 39, and references therein). Alternatively, some guide snoRNPs, in particular those encoded by multicopy genes expressed in brain, can be present in the nucleoplasm, their most-likely site of assembly (40), at concentrations high enough to interact with their targets. Finally, some small RNAs (including U6 snRNA) and mRNAs are known to transit the nucleolus en route to their final destinations (41–43). It will be interesting to look at the morphology of the active multiarray snoRNA loci in neuronal nuclei, to find out whether they are associated with CBs (39) or some other structures func-

tioning as the small nucleolar ribonucle-oprotein (snoRNP) maturation "factory."

The four RNAs discussed above are the first tissue-specific snoRNAs to be described; four additional brain-localized C/D box RNAs have already been identified in rodents (J. Cavaillé, P. Vitali, Z. Basyuk, and J. P. Bachellerie; J. Brosius and A. Hüttenhofer; personal communications). How many more will be found? Do tissues other than brain also have their own sets of snoRNAs? What is the biological role of the newly discovered RNAs and the relevance of their tandemly repeated gene organization? Given the unusual diversity of RNA functions already documented, no doubt many more noncoding RNAs will be found to operate in cells. RNomics approaches, involving either cDNA bank screens similar to the screen used by Cavaillé *et al.* (17) and others (44), or analyses of "intergenic gaps," like that carried out in yeast (45), may be rewarding in identifying additional players in RNA games.

1. Gesteland, R. F., Cech, T. R. & Atkins, J. F., eds. (1999) *The RNA World* (Cold Spring Harbor Lab. Press, Plainview, NY).
2. Eddy, S. R. (1999) *Curr. Opin. Genet. Dev.* **9,** 695–699.
3. Kelley, R. L. & Kuroda, M. I. (2000) *Cell* **103,** 9–12.
4. Karzai, A. W., Roche, E. D. & Sauer, R. T. (2000) *Nat. Struct. Biol.* **7,** 449–455.
5. Burge, C. B., Tuschl, T. & Sharp, P. A. (1999) in *The RNA World*, eds. Gesteland, R. F., Cech, T. R. & Atkins, J. F. (Cold Spring Lab. Press, Plainview, NY), pp. 525–560.
6. Staley, J. P. & Guthrie, C. (1998) *Cell* **92,** 315–326.
7. Estevez, A. M. & Simpson, L. (1999) *Gene* **240,** 247–260.
8. Moss, E. G. (2000) *Curr. Biol.* **10,** R436–R439.
9. Pasquinelli, A. E., Reinhart, B. J., Slack, F., Martindale, M. Q., Kuroda, M. I., Maller, B., Hayward, D. C., Ball, E. E., Degnan, B., Müller, P., *et al.* (2000) *Nature (London)* **408,** 86–89.
10. Bachellerie, J.-P., Cavaillé, J. & Qu, L.-H. (2000) in *The Ribosome: Structure, Function, Antibiotics, and Cellular Interactions*, eds. Garrett, R. A., Douthwaite, S. R., Liljas, A., Matheson, A. T., Moore, P. B. & Noller, H. F. (Am. Soc. Microbiol. Press, Washington, DC), pp. 191–203.
11. Filipowicz, W., Pelczar, P., Pogacic, V. & Dragon, F. (1999) *Acta Biochim. Pol.* **46,** 377–389.
12. Lafontaine, D. L. J. & Tollervey, D. (1998) *Trends Biochem. Sci.* **23,** 383–388.
13. Ofengand, J. & Fournier, M. J. (1998) in *Modification and Editing of RNA*, eds. Grosjean, H. & Benne, R. (Am. Soc. Microbiol. Press, Washington, DC), pp. 229–253.
14. Weinstein, L. & Steitz, J. A. (1999) *Curr. Opin. Cell Biol.* **11,** 378–384.
15. Mitchell, J. R. & Collins, K. (2000) *Mol. Cell* **6,** 361–371.
16. Pogacic, V., Dragon, F. & Filipowicz, W. (2000) *Mol. Cell. Biol.* **20,** 9028–9040.
17. Cavaillé, J., Buiting, K., Kiefmann, M., Lalande, M., Brannan, C. I., Horsthemke, B., Bachellerie, J.-P., Brosius, J. & Hüttenhofer, A. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 14311–14316. (First Published December 5, 2000; 10.1073/pnas.250426397)
18. de Los Santos, T., Schweizer, J., Rees, C. A. & Francke, U. (2000) *Am. J. Hum. Genet.* **67,** 1067–1082.
19. Nicholls, R. D., Saitoh, S. & Horsthemke, B. (1998) *Trends Genet.* **14,** 194–200.
20. Brannan, C. I. & Bartolomei, M. S. (1999) *Curr. Opin. Genet. Dev.* **9,** 164–170.
21. Tilghman, S. M. (1999) *Cell* **96,** 185–193.
22. Nicholls, R. D. (1999) *Nature Genet.* **23,** 132–134.
23. Nicholls, R. D. (2000) *J. Clin. Invest.* **105,** 413–418.
24. Kiss, T. & Filipowicz, W. (1995) *Genes Dev.* **9,** 1411–1424.
25. Tycowski, T. H., You, Z. H., Graham, P. J. & Steitz, J. A. (1998) *Mol. Cell* **2,** 629–638.
26. Ganot, P., Jady, B. E., Bortolin, M.-L., Darzacq, X. & Kiss, T. (1999) *Mol. Cell Biol.* **19,** 6906–6917.
27. Jady, B. E. & Kiss, T. (2000) *Nucleic Acids Res.* **28,** 1348–1354.
28. Cavaillé, J., Nicoloso, M. & Bachellerie, J. P. (1996) *Nature (London)* **383,** 732–735.
29. Omer, A. D., Lowe, T. M., Russell, A. G., Ebhardt, H., Eddy, S. R. & Dennis, P. P. (2000) *Science* **288,** 517–522.
30. Gaspin, C., Cavaillé, J., Erauso, G. & Bachellerie, J.-P. (2000) *J. Mol. Biol.* **297,** 895–906.
31. Burns, C. M., Chu, H., Rueter, S. M., Hutchinson, L. K., Canton, H., Sanders-Bush, E. & Emeson, R. B. (1997) *Nature (London)* **387,** 303–308.
32. Niswender, C. M., Copeland, S. C., Herrick-Davis, K., Emeson, R. B. & Sanders-Bush, E. (1999) *J. Biol. Chem.* **274,** 9472–9478.
33. Yi-Brunozzi, H. Y., Easterwood, L. M., Kamilar, G. M. & Beal, P. A. (1999) *Nucleic Acids Res.* **27,** 2912–2917.
34. Paul, M. S. & Bass, B. L. (1998) *EMBO J.* **17,** 1120–1127.
35. Chen, C.-X., Cho, D.-S. C., Wang, Q., Lai, F., Carter, K. C. & Nishikura, K. (2000) *RNA* **6,** 755–767.
36. Bokar, J. A. & Rottman, F. M. (1998) in *Modification and Editing of RNA*, eds. Grosjean, H. & Benne R. (Am. Soc. Microbiol. Press, Washington, DC), pp. 183–200.
37. Moore, M. J. & Sharp, P. A. (1992) *Science* **256,** 992–997.
38. Bardwell, V. J., Wickens, M., Bienroth, S., Keller, W., Sproat, B. S. & Lamond, A. I. (1991) *Cell* **65,** 125–133.
39. Matera, A. G. (1999) *Trends Cell Biol.* **9,** 302–317.
40. Samarsky, D. A., Fournier, M. J., Singer, R. H. & Bertrand, E. (1998) *EMBO J.* **13,** 3747–3757.
41. Carmo-Fonseca, M., Mendes-Soares, L. & Campos, I. (2000) *Nat. Cell Biol.* **2,** E107–E112.
42. Olson, M. O. J., Dundr, M. & Szebeni, A. (2000) *Trends Cell Biol.* **10,** 189–196.
43. Pederson, T. & Politz, J. C. (2000) *J. Cell Biol.* **148,** 1091–1095.
44. Kiss-László, Z., Henry, Y., Bachellerie, J.-P., Caizergues-Ferrer, M. & Kiss. T. (1996) *Cell* **85,** 1077–1088.
45. Olivas, W. M., Muhlrad, D. & Parker R. (1997) *Nucleic Acids Res.* **22,** 4619–4625.

COMMENTARY